STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **15**, February 2026, pp 1248–1257.
Published online in International Academic Press (www.IAPress.org)

Analyzing Performance Discrepancies: U-Net vs TransUNet for Aircraft Emergency Landing Site Detection

Adil Illi [1,*], El Hadaj Salah [2], Bouzaachane Khadija [1], El Guarmah El Mahdi [3]

[1] *L2IS, Department of Computer Sciences, Faculty of Sciences and Technology, University Cadi Ayyad, Marrakech, Morocco*
[2] *L2IS, National School of Business and Management, University Cadi Ayyad, Marrakech, Morocco*
[3] *L2IS, Mathematics and Informatics Department, Royal Air School, Marrakech, Morocco*

Abstract In the context of aviation, forced landings are unwanted events that can happen to an aircraft during its flight trajectory. They can be due to engine malfunctions, adverse weather conditions and other sudden situations. For this reason, and to ensure passengers' safety, it is imperative to develop methods and procedures to detect potential sites that can be used as emergency landing areas during these crisis situations. Traditionally, pilots use visual indicators to detect such landing sites, this ability can varry from a pilot to another depending on experience, aircraft altitude and other environmental conditions. Such circumstances can make this visual detection task highly difficult.
Image segmentation is one of the possible solutions that can be implemented in identifying potential emergency landing sites for aircraft. Precise segmentation should improve on the effective identification of safe landing areas, thereby enhancing aviation safety protocols in general.
In this context, the traditional U-Net [12] architecture has shown exceptional results regarding segmentation tasks. However, a new approach derived from U-Net and incorporating transformers [13, 4] in its encoder, known as TransUNet [3], has demonstrated promising results, surpassing in some cases those of U-Net.
This study investigates the performance of TransUNet compared to traditional U-Net for aircraft emergency landing site detection. Both architectures were implemented, trained, and evaluated using our novel dataset tailored for this purpose. Our work demonstrate that U-Net outperforms TransUNet in terms of accuracy and computational efficiency in this specific segmentation task. In particular, U-Net exhibited superior performance by improving segmentation precision from 80% up to 88% in the testing set. Moreover, the mean Intersection-Over-Union, a metric for segmentation accuracy, have also seen an improvement of 77% for U-Net over 73% for TransUNet. These results emphasise the power of the traditional U-Net architecture for this critical application, underlying its practical relevance in enhancing aviation safety.

Keywords Vision Transformer, Unet, TransUnet, Emergency Landing Site Detection, Semantic Segmentation.

DOI: 10.19139/soic-2310-5070-2753

1. Introduction

The increasing frequency of air travel necessitates robust emergency preparedness to ensure the safety of passengers and crew. In critical situations, the ability to identify suitable emergency landing sites promptly can significantly reduce risks. Automating the segmentation of potential emergency aircraft landing sites from aerial imagery allows for real-time, precise decision-making, minimizing the reliance on manual assessments and enhancing the speed of response. In this research, we evaluate two leading deep learning architectures, U-Net and TransUnet, both recognized for their strong performance in image segmentation. Our objective is to compare these models to determine which is better suited for the critical challenge of accurately identifying safe landing areas across diverse terrains and under varying environmental conditions.

*Correspondence to: Adil ILLI (Email: a.illi.ced@uca.ac.ma) L2IS, University Cadi Ayyad, Marrakech, Morocco

ISSN 2310-5070 (online) ISSN 2311-004X (print)

This paper first lays the groundwork in Section 2 with a review of relevant literature. We then describe our experimental design in Section 3, covering the architectures, data, and metrics. We conclude in Section 4 with a presentation and detailed analysis of our results.

2. Background and Related Work

Classical images have been used in previous works to identify safe landing zones for aircraft and UAVs. Mejias et al. [11] used a multiclass support vector machine (SVM) approach, while Warren et al. [14] compared the results of applying a Canny Edge detector to a 2D image with the effect of 3D restoration using the Structure-from-Motion method.

Considerable attention has also been paid to UAV landing/crash detection. Kikumoto et al. [9] proposed a method for safety heat mapping using Convolutional Neural Networks (CNNs) and optical flow analysis. Hinzmann et al. [6] used Canny edge detector and Random Forest classifier to classify the landscape.

In other ways, Eendebak et al. [5] developed a real-time emergency landing system using background computation and remote sensing modeling, and Kalzahi et al. [1] used the Gabor Transform and Markov chain rules to predict potential landing sites.

Recent research has focused on CNN-based semantic segmentation techniques, such as the U-Net architecture combined with temporary dense connect modules [8] but some methods, such as the UAV crash management system Safe2Ditch[10], do not rely on image processing techniques but use a pre-determined database of potential crash sites.

The current methodologies for aircraft emergency landing area detection, particularly for UAVs, face several significant limitations. Firstly, the focus on single point landing for UAVs constrains the adaptability of emergency landing procedures, which ideally should consider continuous rectangular shaped landing sites to account for long landing distances. Secondly, there is a notable absence of a dedicated segmentation dataset specifically for safe landing areas. This lack hinders the development and fine-tuning of models tailored to accurately identify and segment potential landing zones in diverse environments. The effectiveness of existing models is often compromised by their reliance on less relevant training data, hindering their performance in real-world scenarios. Furthermore, many current approaches overlook the advantages of state-of-the-art segmentation methods, such as transformer-based models, which have proven superior in other demanding image analysis tasks. This reluctance to adopt advanced techniques results in models that lack the accuracy and robustness essential for reliability in critical emergency situations. Addressing these shortcomings is therefore a crucial step toward building safer and more capable emergency landing systems for UAVs.

This study evaluates the U-Net [12] and TransUNet [3] architectures for segmenting safe landing areas in aerial imagery. Although both models are prominent in medical imaging, a direct comparison for this application represents a significant research gap. Our work is motivated by findings in the medical field where TransUNet's transformer-based design often gives it a performance edge over U-Net. Therefore, we seek to verify whether this performance advantage is maintained when these architectures are applied to the complex task of identifying safe landing zones from above.

3. Proposed Methodology

This section details the methodology we designed to evaluate the selected models. We will first give a brief introduction of the Unet and TransUnet architectures, then present the dataset that was used for this study and detail the evaluation metrics that were implemented to evaluate the performance of both models.

3.1. Models

For this study, we have chosen the Unet and TransUnet architectures for multiple reasons. Both models have demonstrated great potential in various image processing fields. In particular, in medical imaging, the models

exhibited exceptional performance. In fact, a previous study [2] compared these models in the context of medical images and demonstrated the superiority of TransUnet. Through our study, we examine whether this superiority holds in the task of segmenting emergency aircraft landing areas.

- **Unet** is a type of convolutional neural network with two main parts: a contracting path and an expansive path. The contracting path works like a standard convolutional network, gradually shrinking the data. On the other hand, the expansive path grows the data back up, with each step involving upsampling the feature map and then passing it through an up-convolution layer.

The U-shaped structure of UNet (see figure 1) comprises a succession of contraction layers followed by expansion layers, allowing for the progressive capture of features at different spatial scales. The skip connections between the contraction and expansion layers facilitate the merging of information at different resolutions for precise segmentation.

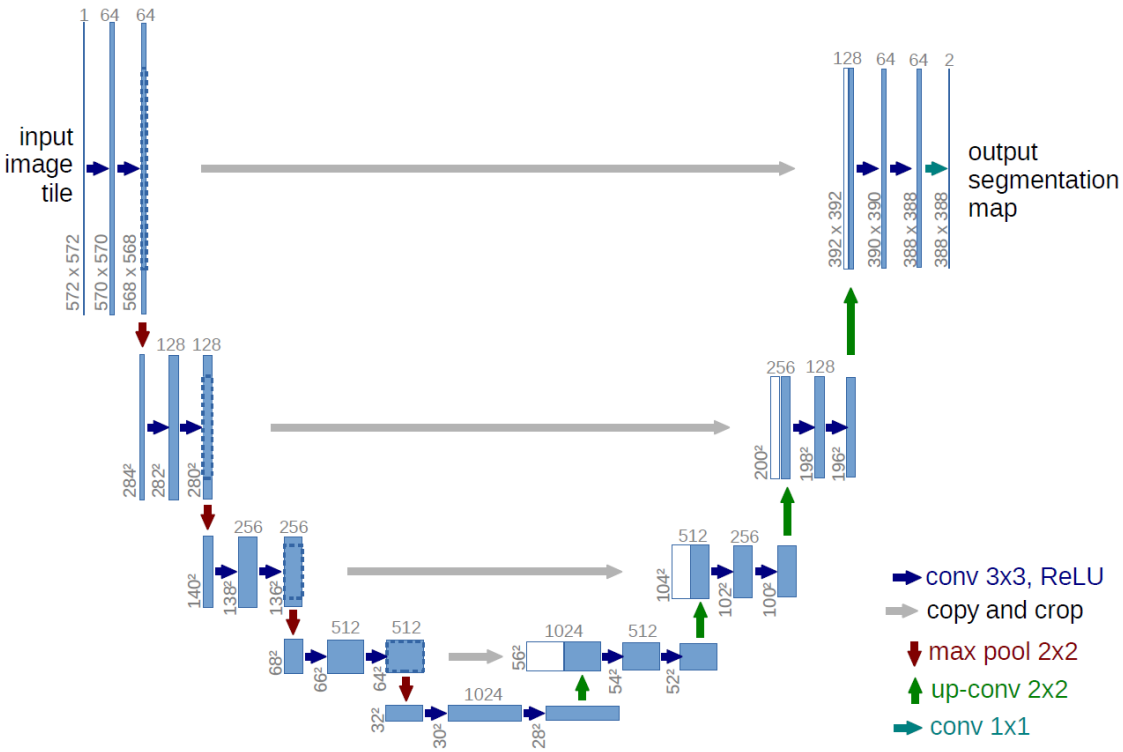

Figure 1. Unet model architecture [12].

- **TransUnet** blends the strengths of Transformers and U-Net. The Transformer takes image patches from a convolutional neural network (CNN) feature map, turning them into a sequence to capture global context. Meanwhile, the decoder expands these encoded features and merges them with high-resolution CNN feature maps to pinpoint precise locations.

The architecture of TransUnet (see figure 2) follows a hybrid structure, merging vision transformer blocks with U-Net contraction-expansion blocks, thereby harnessing the power of attention and the detail-capturing capability of U-Net for precise segmentation at different scales.

3.2. Data description

We used our own dataset [7] of manually annotated images (see sample in figure 3) for the segmentation of emergency aircraft landing sites to distinguish between two main categories: safe and unsafe landing sites. Through manual annotation, each image was labeled to identify areas considered safe for emergency landing from

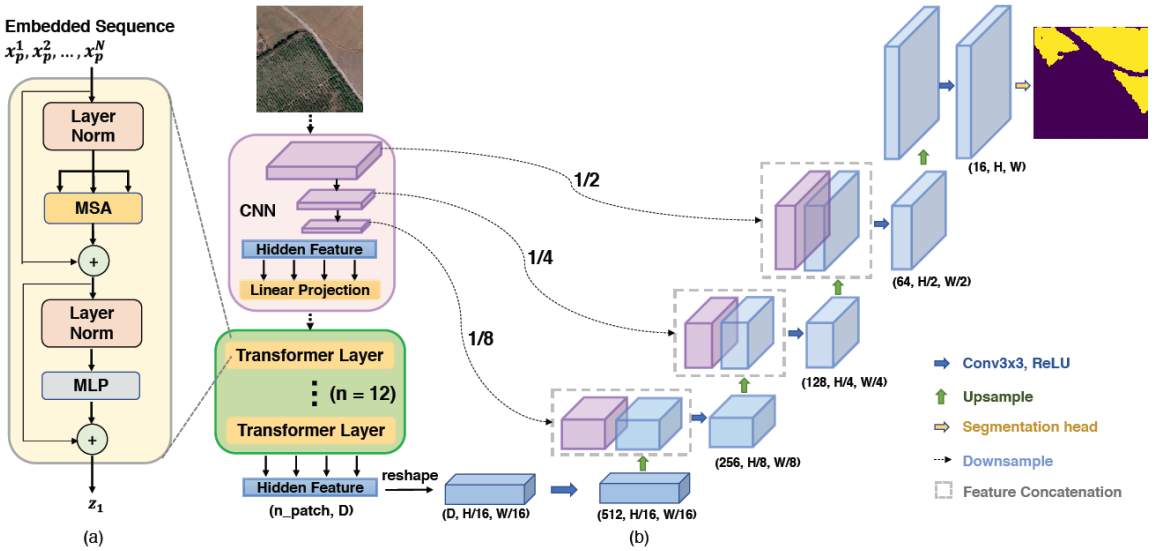

Figure 2. TransUnet model architecture [3].

potentially hazardous areas. This large dataset provides invaluable ground truth data for training and evaluating machine learning models aimed at automatically estimating aircraft landing areas during emergency situations. This dataset contains 4180 images divided into raw google map image and annotation masks. The dataset provides comprehensive coverage across a diverse range of environments, including urban, rural, forest, grassland, and mountainous terrains. We acknowledge that the exclusion of snow-covered landscapes represents a limitation of this study.

Figure 3. Sample of the dataset used [7].

We employed a comprehensive data augmentation strategy to improve model robustness and generalization. During training, we applied random transformations—including rotations, flips, scaling, and brightness/contrast adjustments—to simulate diverse, real-world viewing conditions. This process ensures the model learns to effectively manage the variability inherent in aerial imagery.

3.3. Evaluation metrics

Given the critical safety implications of this task, the choice of evaluation metrics is essential for validating a model's effectiveness. Accordingly, we assessed the performance of our segmentation algorithms against ground truth annotations using the following metrics:

- **Mean Intersection over Union (mIoU)** is a standard metric for judging the accuracy of a segmentation model. It measures how well the predicted segmentation mask (A) overlaps with the actual ground truth mask (B). Conceptually, it calculates the ratio of the correctly identified area (the intersection of the two masks) to the total area covered by both masks combined (their union).

$$IoU = \frac{A \cap B}{A \cup B} \tag{1}$$

- **Dice similarity coefficient (DSC)** measures the overlap or similarity between two sets. In the context of image segmentation it is frequently used when evaluating the similarity of two binary masks, for example a ground truth mask A and a prediction mask B.

$$DSC = 2 \times \frac{|A \cap B|}{|A| + |B|} \tag{2}$$

where $|.|$ denotes the cardinality of a set

- **Accuracy** provides a straightforward measure of how well a model performs in the segmentation task. it represents the ratio of correctly predicted instances to the total number of instances in the dataset.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

Where:
TP = True Positives
TN = True Negatives
FP = False Positives
FN = False Negatives

- **Precision** evaluates the degree to which a model recognizes relevant pixels. It is the proportion of accurately separated true positive pixels to all pixels that the algorithm recognized as positive including the false positives. Higher precision means that the method tends to be accurate when it classifies a pixel as associated with the target class.

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

- **Recall** measures a model's ability to precisely identify all relevant pixels in an image. It is sometimes referred to as sensitivity or true positive rate. It is the proportion of true positive pixels to all the pixels that make up the real target object including the false negatives. A higher recall value means that the model returns most of the relevant results.

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

- **False Positive Ratio (FPR)** quantifies the proportion of actual negative cases that are incorrectly classified as positive. A high FPR means the segmentation model frequently mislabels unsafe areas as safe, which can be hazardous by suggesting unsuitable landing zones. A low FPR indicates that the model rarely produces such false alarms, contributing to safer landing site identification.

$$FPR = \frac{FP}{FP + TN} \tag{6}$$

These metrics provide a comprehensive view of segmentation success, enabling a thorough performance analysis of the two models. By examining recall, precision, accuracy, DSC, IoU and FPR, we can identify the strengths and weaknesses of Unet and TransUnet in the specific context of emergency aircraft landing area segmentation. This detailed evaluation helps in understanding not only which model performs better overall but also in what specific aspects one model may outperform the other.

4. Experimental results and discussion

4.1. Training

We trained both models using the same dataset, previously introduced, during training we noticed that Unet reachs the minimum validation loss much faster than TransUnet. This is due to it's ability to generalize well even with limited data. Table 1 details the training parameters implemented for both models.

Table 1. Unet and TransUnet training parameters.

Model	**Unet**	**TransUnet**
Number of parameters	6,667,639	91,718,993
Input image size	256×256	
Optimizer	adam	
Learning Rate	0.01	
Batch size	8	
Max epochs	100	

In Figure 4 below, we present the loss history graphs for the training and validation subsets. We observe that Unet reaches the lowest loss value of 0.26 faster than TransUnet, where the lowest loss value is 0.48. Both models show signs of overfitting after these points. To mitigate this, we applied TensorFlow's "save best only" option in the ModelCheckpoint callback, which stores the model weights corresponding to the lowest validation loss encountered during training, rather than the final epoch. This approach does not constitute early stopping; instead, it ensures that evaluation is performed on the model state with the best generalization performance, thus preventing the degradation caused by overfitting.

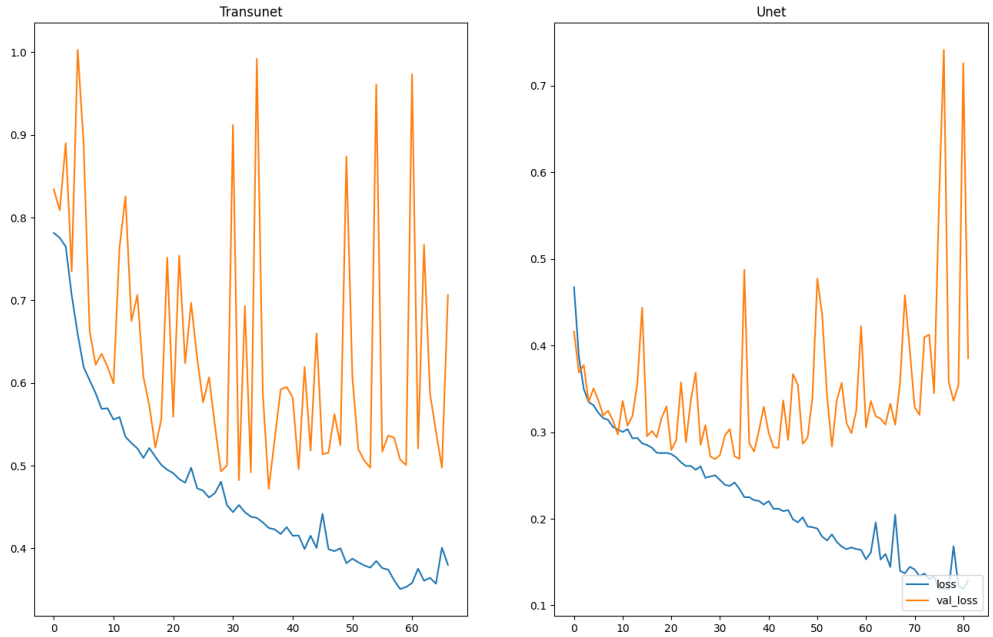

Figure 4. Training loss for testing and validation subsets.

4.2. Evaluation

To start our analysis, we created confusion matrices for both models, shown in Figure 5 below. These matrices break down each model's performance by detailing the counts of true positives, true negatives, false positives, and false negatives. Looking at these matrices helps us understand what each model does well and where it struggles in making accurate predictions.

An analysis of the confusion matrices reveals a nuanced, class-dependent performance difference between the models. The UNet architecture demonstrates a clear and significant advantage in the classification of safe pixels. This is evidenced by its higher true positive and true negative rates, which suggests a more robust feature-learning mechanism for identifying viable landing areas.

Conversely, this superiority does not extend to the detection of unsafe pixels. In this more challenging context, both models perform comparably, with neither showing a distinct edge. This performance parity suggests that identifying the complex features associated with hazardous terrain is an inherent difficulty for both architectures.

This analysis underscores the importance of moving beyond global accuracy metrics. While UNet is demonstrably more reliable for confirming safe zones, the shared struggle to detect unsafe areas highlights a critical avenue for future research. Understanding these specific strengths and weaknesses is paramount for developing truly trustworthy systems for safety-critical applications.

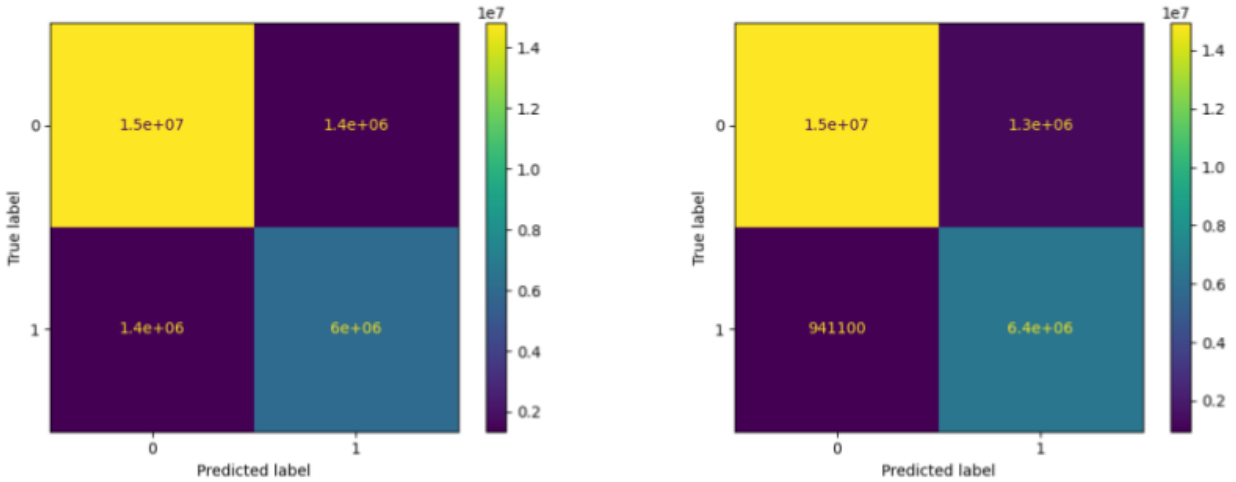

Figure 5. Confusion matrices for TransUnet (on the left) and Unet model (on the right).

Then we calculated the metrics' values for the testing subset for both model, as shown in table 2. While both gave acceptable results, Unet clearly surpasses TransUnet on all our selected metrics. Given that our dataset is relatively small, UNet perform better because of its ability to generalize well with less information. UNet's superior performance may be explained by TransUNet's incorporation of transformer-based attention mechanisms, which aren't always necessary or beneficial, especially for tasks with simple structures or limited spatial dependencies. For these kind of tasks, UNet's architecture which relies on skip connections and feature concatenation seems more adequate.

In addition to segmentation accuracy, we evaluated the inference time of both models to assess their suitability for real-time and embedded applications. The results indicate that U-Net achieves an average inference time of 3 ms per image, compared to 26 ms per image for TransUNet, demonstrating significantly lower latency. This improvement can be explained by U-Net's smaller parameter count and reduced computational complexity, which make it more efficient for deployment on resource-constrained platforms. While training time and hardware-specific optimizations were not the focus of this study, they represent important avenues for future investigation to further assess operational feasibility.

Table 2. Evaluation metrics measurement for Unet and TransUnet.

Model	**Unet**	**TransUnet**
Accuracy	90.66%	90.29%
Precision	83.64%	82.38%
Recall	87.27%	87.00%
mIoU	80.84%	80.27%
DSC	85.41%	85.00%
FPR	7.79%	8.56%

4.3. Results

Finally, we present some test examples for segmentation using both models. These examples further confirm that UNet is the best candidate for detecting the safe landing zones. In fact, UNet is better at predicting borderline cases that may lead dangerous situations if they are marked as safe. In the last example of Figure 6, TransUNet detected an unsafe area as safe, which does not align with the ground truth. This could be very dangerous in real-life scenarios. However, UNet accurately segmented this boundary as unsafe for landing.

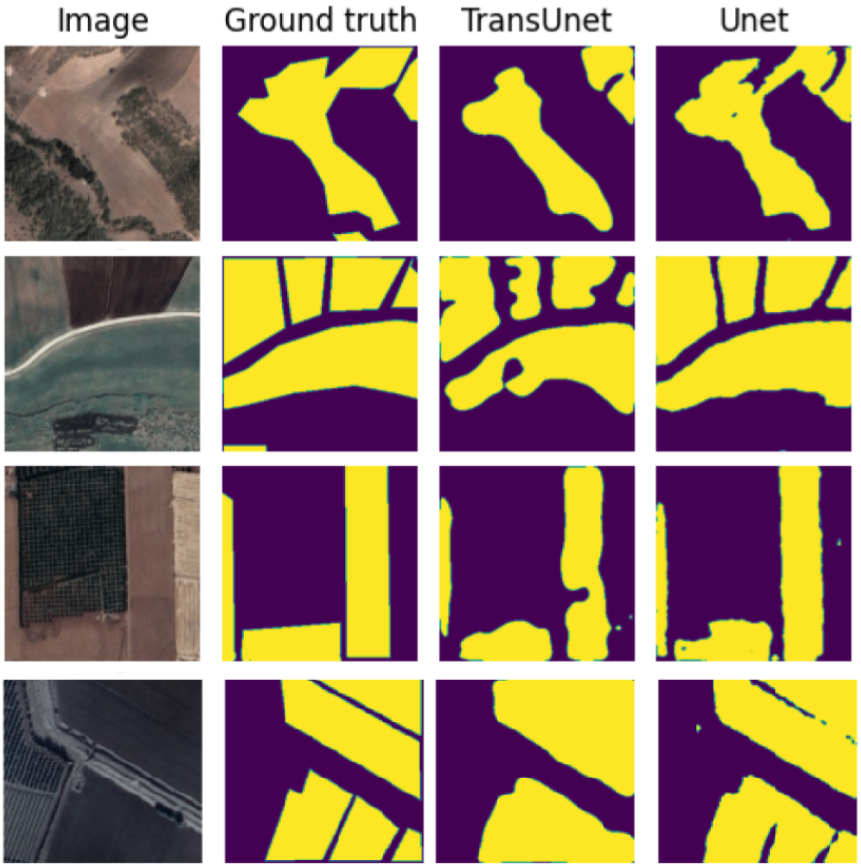

Figure 6. Examples of testing images and models' prediction.

4.4. Adverse Weather Simulation Results

To evaluate robustness, we tested both architectures under light fog, heavy fog, and rain simulations, conditions that commonly degrade visibility in real-world flight scenarios. The adverse conditions were generated using the publicly available framework by Yang [15], which provides realistic weather overlays for computer vision tasks. Both models experienced a notable performance drop compared to clear-weather data. For U-Net, mIoU decreased from 79.28% (raw) to 61.26% in light fog, 13.55% in heavy fog, and 18.38% in rain, with accuracy falling to 76.24%, 25.37%, and 32.87%, respectively. U-Net's recall remained relatively high at 96.13% (light fog), 65.40% (heavy fog), and 80.15% (rain), but false positive ratios increased significantly to 33.82%, 94.88%, and 91.05%, respectively. TransUNet followed a similar trend, with mIoU dropping from 80.20% (raw) to 56.42% in light fog, 16.83% in heavy fog, and 16.90% in rain, with accuracies of 72.24%, 33.63%, and 33.70%, respectively. TransUNet maintained near-perfect recall (97.66% in light fog, 99.99% in heavy fog, and 99.96% in rain) but at the cost of extremely high false positive ratios (40.63%, 99.94%, and 99.82%, respectively), making its predictions unsafe for operational use. U-Net, while also affected, showed a more balanced trade-off across all conditions.

To visually illustrate these findings, we present Figure X, which shows representative examples from the testing dataset under light fog, heavy fog, and rain simulations. Each row displays the raw image, its weather-simulated counterpart, and the corresponding segmentation outputs from U-Net and TransUNet. Under all conditions, TransUNet tends to over-segment safe areas, frequently misclassifying unsafe regions as safe, particularly in heavy fog and rain. U-Net, despite reduced accuracy, demonstrates more conservative and reliable predictions across all weather conditions. These qualitative results reinforce the quantitative findings, highlighting the significant vulnerability of both models to weather artifacts, with U-Net exhibiting comparatively more robust behavior, especially in light fog.

5. Conclusion

While our results indicate that U-Net currently outperforms TransUNet in segmenting emergency aircraft landing sites, transformer-based architectures still hold significant potential for more complex or large-scale scenarios. U-Net's encoder-decoder structure effectively captures the fine-grained spatial details necessary for identifying safe landing zones. TransUNet's underperformance in our study may be attributed to factors such as limited pretraining and insufficient data for fully leveraging attention mechanisms. We also note that aircraft-specific constraints, such as minimum runway length, slope, and obstacle presence, were not incorporated in the current evaluation, representing a limitation of this study. Future work should explore enhanced pretraining, larger and more diverse datasets, multi-modal inputs, and real-time optimization to unlock the full capabilities of transformer-based models for emergency landing site detection. The conclusion has been adjusted in the revised manuscript to reflect these nuances. Future work could explore the integration of advanced transformer models, multi-modal data, and real-time optimization techniques to enhance segmentation accuracy and deployability for emergency landing site detection.

Acknowledgment

The authors express their heartfelt gratitude to the Computer Science Department, the Laboratory of Computer Engineering and Systems (L2IS) at the Faculty of Science and Technology (FST), the National School of Business and Management, and the Royal Air School of Aeronautics for generously providing the essential facilities and resources that supported this study.

REFERENCES

1. M. Asadzadeh Kaljahi, P. Shivakumara, M. Yamani Idna Idris, M. Hossein Anisi, T. Lu, M. Blumenstein, and N. Mohamed Noor. An automatic zone detection system for safe landing of uavs. Expert Systems with Applications, vol 122, pp 319-333 (2019). doi: 10.1016/j.eswa.2019.01.024
2. R. Castro, L. Ramos, S. Román, M. Bermeo, A. Crespo and E. Cuenca. U-Net vs. TransUNet: Performance Comparison in Medical Image Segmentation. Applied Technologies. Communications in Computer and Information Science, vol 1755, pp 212–226 (2023). doi: 10.1007/978-3-031-24985-3_16
3. J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille and Y. Zhou. TransUNet: transformers make strong encoders for medical image segmentation. Computing Research Repository, pp 1-13 (2021). doi: 10.48550/arXiv.2102.04306
4. A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit and N. Houlsby, An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. International Conference on Learning Representations (2021). doi: 10.48550/arXiv.2010.11929
5. P. T. Eendebak, A. W. M. van Eekeren, and R. J. M. den Hollander. Landing spot selection for uav emergency landing. Defense, Security, and Sensing, vol 8741, pp 87410G (2013). doi: 10.1117/12.2017910
6. T. Hinzmann, T. Stastny, C. Cadena, R. Siegwart, and I. Gilitschenski. Prior-free visual landing site detection for autonomous planes. IEEE International Conference on Robotics and Automation, vol 3, no 3, pp 2521-2528 (2018). doi: 10.1109/LRA.2018.2809962
7. A. Illi, K. Bouzaachane, S. El Hadaj and E.M. El Guarmah, A pixel-wise labelled dataset of Moroccan aircraft emergency landing sites for semantic segmentation applications, Data In Brief, vol 54, 110379 (2024). doi: 10.1016/j.dib.2024.110379
8. B. Jiang, Z. Chen, J. Tan, R. Qu, C. Li, and Y. Li. A real-time semantic segmentation method based on STDC-CT for recognizing uav emergency landing zones. Sensors, vol 23, no 14, article 6514 (2023). doi: 10.3390/s23146514
9. C. Kikumoto, Y. Harimoto, K. Isogaya, T. Yoshida and T. Urakubo. Landing site detection for uavs based on cnns classification and optical flow from monocular camera images. Journal of Robotics and Mechatronics, vol 33, pp 292-300, (2021). doi: 10.20965/jrm.2021.p0292
10. P. C. Lusk, P. C. Glaab, L. J. Glaab, and R. Beard. Safe2ditch: Emergency landing for small unmanned aircraft systems. Computing, Information and Communication 16(1):1-13 (2019). doi: 10.2514/1.I010706
11. L. Mejias. Classifying natural aerial scenery for autonomous aircraft emergency landing. International Conference on Unmanned Aircraft Systems, 1236-1242, (2014). doi: 10.1109/ICUAS.2014.6842380
12. O. Ronneberger, P. Fischer and T. Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention, pp 234–241, (2015). doi: 10.1007/978-3-319-24574-4_28
13. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser and I. Polosukhin, Attention Is All You Need (2023). doi: 10.48550/arXiv.1706.03762
14. M. Warren , L. Mejias, Y. Xilin, B. Arain, L. Gonzalez and B. Upcroft. Enabling Aircraft Emergency Landings Using Active Visual Site Detection, Springer Tracts in Advanced Robotics, vol 105, pp 167–181 (2015). doi:10.1007/978-3-319-07488-7_12
15. Y. Runou. Adverse Weather Simulation, 2023. url:https://github.com/RicardooYoung/AdverseWeatherSimulation